NFORMATION AND COMPUTATION **81**, 265–279 (1989)

Church–Rosser Controlled Rewriting Systems and Equivalence Problems for Deterministic Context-Free Languages

G. SÉNIZERGUES

LITP et Université de Paris 7, 2 Place Jussieu, 75251-Paris Cedex 05, France

We use a description of deterministic context-free languages (dcfl) by a special type of Church–Rosser rewriting systems to extend several procedures testing the equality of two languages in a subfamily F of dcfl to procedures testing the equality of two languages one of which is in F, the other beeing a general dcfl.

I. INTRODUCTION

We deal here with the following problem: given two families of languages F and C (which can be defined by some special kinds of automata, grammars, or rewriting systems) can we decide for every pair of languages (L, L') in $F \times C$, whether $L = L'$? We shall denote this problem $\mathrm{Eq}(F, C)$.

Let D be the family of deterministic context-free languages (dcfl for short). Since the year 1966 (Ginsburg and Greibach, 1966), many attempts have been made to solve $\mathrm{Eq}(D, D)$. Though this problem remains open, $\mathrm{Eq}(F, D)$ has been shown decidable for some subfamilies F of the family D. Namely for the families

$F =$ the family of finite-turn-dcfl (FT)

$F =$ the family of one-counter dcfl (C)

$F =$ the family of strict-real-time-dcfl (R_0)

$F =$ the family of real-time dcfl (R)

algorithms have been designed to solve $\mathrm{Eq}(F, D)$ (see Oyamaguchi, 1987; Oyamaguchi *et al.*, 1980, 1981). (M. Oyamaguchi noticed that $C \subset R$ (Oya 1) and it is clear that $R_0 \subset R$.)

The algorithms cited above use descriptions of languages in terms of automata. Here we shall also use a description of dcfl over an alphabet X in terms of rewriting systems over the same alphabet X. This technique was

0890-5401/89 $3.00

already used in (Butzbach, 1978). It is shown in Chottin (1979, 1982), that every dcfl is of the form

$$(R)\overset{*}{\underset{S}{\longleftrightarrow}}, \tag{1}$$

where R is a rational set and S is a Church–Rosser controlled rewriting system (see Definition II.1 below). Let us denote by CR the family of languages of the form (1) (so $D \subset CR$).

In part III.A we focus our attention on the family of NTS languages. A language L is NTS (this is an abbreviation for the non-terminal separation property) iff there exists a context-free grammar $G = \langle X, V, P \rangle$ and a set of axioms $A \subset V$ such that $L = L(G, A)$ and for every $v \in V$, $\hat{L}(G, v)$ (the set of sentential forms derived from v) is equal to the class of v for the congruence generated over $(X \cup V)$ by the productions of the grammar. This notion is more precisely defined in Definition III.3 below.

This type of languages is studied in (Autebert, Boasson, and Sénizergues, 1984; Boasson, 1980; Boasson and Sénizergues, 1985; Frougny, 1980; Sénizergues, 1981, 1985). The family of NTS languages (which we shall denote by NTS) is a subfamily of D (Boasson, 1980; Boasson and Sénizergues, 1985; Frougny, 1980).

It has been shown that $\mathrm{Eq}(NTS, NTS)$ is decidable (Sénizergues, 1981, 1985). We state here the

THEOREM III.2. *$\mathrm{Eq}(NTS, D)$ is decidable.*

We recall that the families NTS, FT, R are pairwise incomparable.

In part III.B we prove

THEOREM III.3. *If F is an effective cylinder (i.e., a family of languages effectively closed under inverse homomorphisms and intersection with rational sets) that contains RAT (the family of rational sets) and such that $\mathrm{Eq}(F, F)$ is decidable then $\mathrm{Eq}(F, CR)$ is decidable.*

This theorem gives a new scheme permitting us to extend every decision procedure for $\mathrm{Eq}(F, F)$ to one for $\mathrm{Eq}(F, D)$ in the case where $F = FT$ or $F = C$ or $F = R_0$ or $F = R$.

In part III.C we point out the fact that CR properly contains the family D. Namely, it contains some context-free languages which are not deterministic and also some noncontext-free languages. Hence our scheme seems to give more results than the schemes given in Oyamaguchi *et al.* (1980, 1981).

In part IV we discuss the applicability of our scheme to other families of languages and other types of rewriting systems.

II. Preliminaries

Definition II.1 (Chottin, 1979, 1982). A controlled rewriting system over X is a finite set of triples

$$S = \{(R_i, u_i, v_i) \mid i \in (1, n)\}$$

such that for every integer i, R_i is a regular set over X and u_i, v_i are two words over X such that $|u_i| < |v_i|$.

The direct reduction generated by S is noted $\vdash_S$ and defined by

$$f \underset{S}{\vdash} g \qquad \text{iff} \quad f = rv_is,\ g = ru_is, \text{ where } s \in X^*, i \in [1, n], r \in R_i$$

and (R_i, u_i, v_i) belongs to S.

The reduction generated by S, denoted $\overset{*}{\vdash}_S$, is the reflexive and transitive closure of $\vdash_S$. We denote $\leftrightarrow_S$ the symmetric closure of $\vdash_S$.

The equivalence generated by S, denoted $\overset{*}{\leftrightarrow}_S$, is the reflexive and transitive closure of $\leftrightarrow_S$. One can see that $\overset{*}{\leftrightarrow}_S$ is the finest equivalence relation that contains the set $\{(ru_i, rv_i) \mid i \in [1, n], r \in R_i\}$ and that is a right congruence.

As usual, we say that a word f is irreducible mod(S) iff there exists no g such that $f \vdash_S g$. We denote by Irr(S) the set of irreducible words mod(S). For every controlled rewriting system S over X, Irr(S) is a rational subset of X^*:

$$\mathrm{Irr}(S) = X^* - \bigcup_{i \in [1,n]} R_iv_iX^*. \tag{2}$$

S is said to be Church–Rosser iff for every $f \in X^*$ and $g \in X^*$,

$$f \underset{S}{\overset{*}{\leftrightarrow}} g \qquad \text{iff there exists } w \in X^* \text{ such that } f \underset{S}{\overset{*}{\vdash}} w \text{ and } g \underset{S}{\overset{*}{\vdash}} w.$$

For every word $f \in X^*$, by $[f]\overset{*}{\leftrightarrow}_S$ we denote the class of f modulo $\overset{*}{\leftrightarrow}_S$; that is,

$$[f]\underset{S}{\overset{*}{\leftrightarrow}} = \{g \in X^* \mid f \underset{S}{\overset{*}{\leftrightarrow}} g\}.$$

For every subset $A \subset X^*$, by $[A]\overset{*}{\leftrightarrow}_S$ we denote:

$$[A]\underset{S}{\overset{*}{\leftrightarrow}} = \{g \in X^* \mid \exists a \in A,\ a \underset{S}{\overset{*}{\leftrightarrow}} g\}$$

or equivalently

$$[A]\underset{S}{\overset{*}{\leftrightarrow}} = \bigcup_{a \in A} [a]\underset{S}{\overset{*}{\leftrightarrow}}.$$

We denote by CR the family of Church–Rosser languages which we define by: a language $L \subset X^*$ is in CR iff there exists a Church–Rosser controlled rewriting system S over X and a rational subset R of X^* such that $R \subset \mathrm{Irr}(S)$ and $L = (R) \overset{*}{\leftrightarrow}_S$.

THEOREM II.1 (Chottin, 1979, 1982). *Every deterministic context-free language L can be represented as*

$$L = [R] \underset{S}{\overset{*}{\leftrightarrow}},$$

where S is a Church–Rosser controlled rewriting system and R is a rational set included in Irr(S). *Moreover, the system S and the set R are constructible from a deterministic pushdown automaton recognising L.*

Remark II.1. Theorem II.1 shows that $D \subset CR$.

Remark II.2. — There is in general no algorithm deciding whether a controlled rewriting system S is Church–Rosser or not (this is shown in Proposition IV.1 of Sénizergues, in press 2).

— The Church–Rosser property becomes decidable for some interesting subclasses of controlled rewriting systems (see part IV of Sénizergues, in press 2).

— But, as asserted in Theorem II.1, there is an algorithm producing from every deterministic pushdown automaton a controlled rewriting system S *which is always Church–Rosser* and a rational set $R \subset \mathrm{Irr}(S)$ such that the equality of the theorem is true. A refinement of this theorem giving some additional combinatorial properties of S is shown in Sénizergues (in press 1).

We give now some definitions and results about context-free grammars and especially NTS grammars.

Let $G = \langle X, V, P \rangle$ be a context-free grammar, X is the set of terminal symbols, and V the set of non-terminals. P is the set of productions of G. As usual, by $\to_G$ we denote the direct derivation and by $\overset{*}{\to}_G$ we denote the derivation generated by G. Let us set

$$f \underset{G}{\longmapsto} g \quad \text{iff} \quad g \underset{G}{\to} f$$

and

$$f \underset{G}{\overset{*}{\longmapsto}} g \quad \text{iff} \quad g \underset{G}{\overset{*}{\to}} f:$$

$\overset{*}{\to}_G$ is the relation generated by the semi-Thue system P, while $\overset{*}{\longmapsto}_G$ is the relation generated by the semi-Thue system $P^{-1} = \{(u, v) \mid (v, u) \in P\}$.

$\vdash_G$ is called the direct reduction and $\overset{*}{\vdash}_G$ the reduction associated to G.

$\leftrightarrow_G$ is the union of the direct derivation and of the direct reduction.

We denote by $\overset{*}{\leftrightarrow}_G$ the reflexive and transitive closure of $\leftrightarrow_G$. It is nothing else than the congruence generated by the semi-Thue system P. A word $f \in (X \cup V)^*$ is irreducible mod(G) iff there exists no $g \in (X \cup V)^*$ such that $f \vdash_G g$.

Irr(G) denotes the set of all irreducible words mod(G). For every context-free grammar G, Irr(G) is a rational language over $(X \cup V)^*$, because it can be expressed as

$$\mathrm{Irr}(G) = (X \cup V)^* - (X \cup V)^* \, W (X \cup V)^*, \tag{3}$$

where $W = \{w \in (X \cup V)^*, \exists v \in V \mid (v, w) \in P\}$.

Given a context-free grammar $G = \langle X, V, P \rangle$ and a set of axioms $A \subset (X \cup V)^*$, we define

$$L(G, A) = \{f \in X^* \mid \exists a \in A, a \xrightarrow[G]{*} f\}$$

$$\hat{L}(G, A) = \{f \in (X \cup V)^* \mid \exists a \in A, a \xrightarrow[G]{*} f\}$$

$$[A]_{\overset{*}{\underset{G}{\leftrightarrow}}} = \{f \in (X \cup V)^* \mid \exists a \in A, a \overset{*}{\underset{G}{\leftrightarrow}} f\}$$

$$\Delta(A) = \{f \in (X \cup V)^* \mid \exists a \in A, a \overset{*}{\underset{G}{\vdash}} f\}; \delta(A) = \Delta(A) \cap \mathrm{Irr}(G).$$

Definition II.2 (Boasson, 1980). A context-free grammar $G = \langle X, V, P \rangle$ is said to be NTS iff, for every non-terminal symbol $v \in V$,

$$\hat{L}(G, v) = [v]_{\overset{*}{\underset{G}{\leftrightarrow}}}.$$

Definition II.3 (Boasson, 1980). A language L over X is said to be NTS iff there exists some NTS grammar $G = \langle X, V, P \rangle$ and a finite set of axioms $A \subset V$ such that $L = L(G, A)$.

It is known that:

— given a context-free grammar G, one can decide whether it is NTS or not (Proposition 3 of Sénizergues (1985));

— given a context-free grammar G, one cannot decide whether the language generated by G is NTS or not (Sénizergues, 1981).

A c.f. grammar G is proper iff, $\forall(v, m) \in P$, $m \notin \{\varepsilon\} \cup V$ (that is, no right member of a production is the empty word or a single non-terminal symbol).

We say that G is reduced iff there is no non-terminal v such that $L(G, v) = \emptyset$.

For every NTS grammar G there exists some reduced proper NTS grammar G' generating the same language as G. Moreover, G' is computable from G (Frougny, 1980). Hence, in the following, we only deal with reduced proper NTS grammars.

Remark II.3. Let us define a valuation $\| \ \|$ over $(X \cup V)^*$ by: $\| \ \|$ is the unique homomorphism $(X \cup V)^* \to (\mathbb{N}, +)$ such that

$$\forall v \in V, \qquad \|v\| = 1$$

$$\forall x \in X, \qquad \|x\| = 2.$$

Then, for every proper c.f. grammar G we have

$$f \underset{G}{\longmapsto} g \Leftrightarrow f \underset{G}{\leftrightarrow} g \qquad \text{and} \qquad \|f\| > \|g\|.$$

We recall that NTS languages are deterministic (Boasson, 1980; Boasson and Sénizergues, 1985) and have a decidable equivalence problem (Sen 1, 2). See (Autebert *et al.*, 1984–1, 1984–2; Boasson, 1980; Frougny, 1980; Sénizergues, 1981, 1985) for more details about NTS grammars.

We shall use the following result

Proposition II.1. *If $G = \langle X, V, P \rangle$ is a proper context-free grammar and A a regular set over $X \cup V$ then $\Delta(A)$ and $\delta(A)$ are regular too.*

This is a consequence of Theorem 2–5 of (Book *et al.*, 1982) because P is a monadic semi-Thue system. It is also a consequence of Theorem IV.1 of Sénizergues (1981) because every monadic semi-Thue system is basic. (The first result of this kind was given in Benois, 1969, in the context of the Dyck reduction. For left-basic semi-Thue systems see Sakarovitch, 1979.)

III.A The Extended Equivalence Problem for NTS Languages

We show here that $\mathrm{Eq}(NTS, CR)$ is decidable. Hence $\mathrm{Eq}(NTS, D)$ is decidable.

Let us fix a Church–Rosser language $L = [K] \overset{*}{\longleftrightarrow}_S$, where

$S = \{(R_i, u_i, v_i)\}_{i \in (1,n)}$ is a Church–Rosser controlled rewriting system over X

K is a rational set over X such that $K \subset \mathrm{Irr}(S)$.

We also fix a NTS grammar $G = \langle X, V, P \rangle$, $A \subset V$, and $M = L(G, A)$. We exhibit a procedure to test whether $L = M$ or not. For every pair of languages L_1, L_2 over a same alphabet Y we define the relation $\mathrm{Synt}_Y(L_1, L_2)$ by:

$$\forall (f, g) \in Y^* \times Y^*,\ (f, g) \in \mathrm{Synt}_Y(L_1, L_2)$$

iff

$$\forall u \in L_1, \forall v \in Y^*,\ ufv \in L_2 \Leftrightarrow ugv \in L_2.$$

This generalises the classical notion of syntactic congruence of a language: $\mathrm{Synt}_Y(Y^*, L_2)$ is nothing else than the syntactic congruence of L_2.

PROPOSITION III.1. *$L = M$ iff*

(1) $K \subset M$

(2) $\mathrm{Irr}(S) - K \subset \bar{M}$

(3) $\forall i \in [1, n]$, $(u_i, v_i) \in \mathrm{Synt}_X(R_i, M)$

(*where $\bar{M}$ denotes the complement of M in X^**).

Proof. As S is Church–Rosser, $\bar{L} = [\mathrm{Irr}(S) - K] \overset{*}{\leftrightarrow}_S$. We then have always $K \subset L$ and $\mathrm{Irr}(S) - K \subset \bar{L}$.

Moreover, the relation $\leftrightarrow_S$ saturates L. This is equivalent to: $\forall i \in [1, n]$,

$$(u_i, v_i) \in \mathrm{Synt}_X(R_i, L).$$

Only if. If $L = M$, replacing L by M in the above properties we get properties 1, 2, and 3.

If. Let us suppose that 1, 2, 3 are true:

By 3, $\leftrightarrow_S$ saturates M. It follows that $\leftrightarrow_S$ saturates M.

By 1, $K \subset M$. As $\leftrightarrow_S$ saturates M, we have $[K] \overset{*}{\leftrightarrow}_S \subset M$.

By 2, we get $[\mathrm{Irr}(S) - K] \overset{*}{\leftrightarrow}_S \subset \bar{M}$.

Hence $L \subset M$ and $\bar{L} \subset \bar{M}$ which implies $L = M$. ∎

Conditions 1 and 2 are clearly decidable because M and $\bar{M}$ are deterministic cfl while K and $\mathrm{Irr}(S) - K$ are regular. In order to see that 3 is decidable, we show that for every regular set $R \subset X^*$, $\mathrm{Synt}_X(R, M)$ is a computable relation.

Let us denote $\hat{M} = \hat{L}(G, A)$ ($\hat{M}$ is a language over $X \cup V$).

LEMMA III.1. $\mathrm{Synt}_X(R, M) = \mathrm{Synt}_{X \cup V}(\delta(R), \hat{M}) \cap X^* \times X^*$.

Proof. (1) Let (f, g) belong to $\mathrm{Synt}_X(R, M)$. Let u, v be such that $u \in \delta(R)$, $v \in (X \cup V)^*$, and $ufv \in M$:

There exists $\bar{u} \in R$ such that $\bar{u} \overset{*}{\longmapsto}_G u$.

There exists $\bar{v} \in X^*$ such that $\bar{v} \overset{*}{\longmapsto}_G v$ (because G is reduced).

$ufv \overset{*}{\to}_G \bar{u}f\bar{v}$; hence $\bar{u}f\bar{v} \in M$. But $(f, g) \in \mathrm{Synt}_X(R, M)$; hence $\bar{u}g\bar{v} \in M$, so $ugv \in \hat{M}$.

(2) Let (f, g) belong to $\mathrm{Synt}_{X \cup V}(\delta(R), \hat{M}) \cap X^* \times X^*$. Let $(u, v) \in R \times X^*$ such that $ufv \in M$. As the relation $\longmapsto_G$ strictly reduces the valuation $\| \; \|$, there exists some word $u' \in \mathrm{Irr}(G)$ such that $u \overset{*}{\longmapsto}_G u'$ (this word u' is then in $\delta(R)$):

$ufv \overset{*}{\longmapsto}_G u'fv$, which by the NTS condition implies: $u'fv \in \hat{M}$.

As $u' \in \delta(R)$ and $(f, g) \in \mathrm{Synt}_{X \cup V}(\delta(R), \hat{M})$ we have also: $u'gv \in \hat{M}$.

As $u'gv \overset{*}{\to}_G ugv$ and $ugv \in X^*$, we conclude that $ugv \in M$. ∎

By Proposition II.1, if R is regular then $\delta(R)$ is regular. To state that $\mathrm{Synt}_X(R, M)$ is computable for every regular set R over X, it is then sufficient to prove that $\mathrm{Synt}_{X \cup V}(R, \hat{M})$ is computable for every regular set R over $X \cup V$ such that $R \subset \mathrm{Irr}(G)$.

Let $\#$ be a new letter ($\# \notin X \cup V$). For every language R and word m over $X \cup V$ we define:

$$C(R, M) = \{u \# v \mid u \in R, v \in \mathrm{Irr}(G), \text{ and } umv \in \hat{M}\}.$$

LEMMA III.2. *Let f, g belong to $(X \cup V)^*$:*

$$(f, g) \in \mathrm{Synt}_{X \cup V}(R, \hat{M}) \Leftrightarrow C(R, f) = C(R, g).$$

Proof. (1) Let us suppose that $(f, g) \in \mathrm{Synt}_{X \cup V}(R, \hat{M})$. Let $u \# v \in C(R, f)$. Then $ufv \in \hat{M}$ and $u \in R$. Hence $ugv \in \hat{M}$, so that $u \# v \in C(R, g)$.

We have proved that $C(R, f) \subset C(R, g)$ and the converse can be proved by the same arguments.

(2) Let us suppose that $C(R, f) = C(R, g)$. Let us show that $(f, g) \in \mathrm{Synt}_{X \cup V}(R, \hat{M})$. Let $u \in R$, $v \in (X \cup V)^* \mid ufv \in \hat{M}$. As the relation $\longmapsto_G$ strictly reduces the valuation, there exists some $v' \in \mathrm{Irr}(G)$ such that $v \overset{*}{\longmapsto}_G v'$:

$$ufv \in \hat{M} \quad \text{and} \quad ufv \overset{*}{\underset{G}{\longmapsto}} ufv'.$$

By the NTS condition, $ufv' \in \hat{M}$; hence $u \# v' \in C(R, f)$.

This implies $u \# v' \in C(R, g)$ (because of the hypothesis). Hence $ugv' \in \hat{M}$. $ugv' \overset{*}{\to}_G ugv$, so that $ugv \in \hat{M}$.

By same means we could prove that for every $u \in R$, $v \in (X \cup V)^*$:

$$ugv \in \hat{M} \Rightarrow ufv \in \hat{M}.$$

Hence $(f, g) \in \mathrm{Synt}_{X \cup V}(R, \hat{M})$. ∎

LEMMA III.3. *Let R be a regular language over $X \cup V$, included in* $\mathrm{Irr}(G)$ *and m a word over $X \cup V$. $C(R, m)$ is a one-turn deterministic context-free language.*

Proof. (1) The proof of this lemma is very similar to the proof of Proposition 2, p. 307 of Sénizergues (1985).

The reader should look at the automaton $\mathscr{A}$ described in this article and notice that:

$$L(\mathscr{A}) \subset \{u \# v \mid u, v \in (X \cup V)^* \text{ and umv} \in M\} \tag{1}$$

$$C(R, m) \subset L(\mathscr{A}). \tag{2}$$

Hence $C(R, m) = L(\mathscr{A}) \cap R \# \mathrm{Irr}(G)$ so $C(R, m)$ is one-turn deterministic.

(2) Lemma III.3 is also a particular case of Proposition VI.1, p. 186 of Sénizergues (1987). A complete formal proof can be found in pp. 187–211 of (Sénizergues, 1987).

The equivalence problem for one-turn deterministic pda is decidable (Beeri, 1976; Valiant, 1974). Hence the equality $C(R, f) = C(R, g)$ can be tested and by Lemma III.2, $\mathrm{Synt}_{X \cup V}(R, \hat{M})$ is computable.

We have proved

THEOREM III.1. Eq(NTS, CR) *is decidable.*

By Theorem II.1 we deduce

THEOREM III.2. Eq(NTS, D) *is decidable.*

III.B. OTHER APPLICATIONS OF THE SAME SCHEME

THEOREM III.3. *Let F be a family of languages such that*

(i) *F contains the family RAT of all rational sets*

(ii) *F is effectively closed under inverse homomorphism and intersection with rational sets*

(iii) Eq(F, F) *is decidable.*

Then Eq(F, CR) *is decidable.*

Remark. In the above statement we could replace condition (i) by condition (i′): F contains at least one language L such that $\varepsilon \in L$ (because every cylinder containing such a L contains the whole family RAT).

Proof of the Theorem. Let $L \in CR$ and $M \in F$ be languages over an alphabet X, where

$L = [K] \overset{*}{\longleftrightarrow}_S$

$S = \{(R_i, u_i, v_i)\}_{i \in (1,n)}$ is a Church–Rosser controlled rewriting system and

K is a rational set included in $\mathrm{Irr}(S)$.

Proposition III.1 still applies here:

Condition $1 \Leftrightarrow K = M \cap K$, which is decidable by assumptions (i), (ii), and (iii).

Condition $2 \Leftrightarrow (\mathrm{Irr}(S) - K) \cap M = \varnothing$, which is also decidable by (i), (ii), and (iii).

Condition 3 will be decidable if $\mathrm{Synt}_X(R, M)$ is computable for every rational set R.

Let us define for every $m \in X^*$:

$$LS(R, m) = \{u \# v \mid u \in R, v \in X^* \text{ and } umv \in M\}.$$

FACT 1. $(f, g) \in \mathrm{Synt}_X(R, M) \Leftrightarrow LS(R, f) = LS(R, g)$.

Proof. This fact is a straightforward consequence of the definition of the relation $\mathrm{Synt}_X(R, M)$. ∎

FACT 2. $\forall m \in X^*$, $\forall R \in \mathrm{Rat}(X^*)$, $LS(R, m) = \Phi^{-1}(M) \cap H$, *where* $\Phi: (X \cup \{\#\})^* \to X^*$ *preserves every letter* $x \in X$,

$$\textit{sends } \# \textit{ on} \qquad \Phi(\#) = m$$

and H *is the rational set* $R \# X^*$. *Hence* $LS(R, m) \in F$.

By Fact 1 and Hypothesis (iii), $\mathrm{Synt}_X(R, m)$ is computable. ∎

COROLLARY III.1. $\mathrm{Eq}(FT, CR)$, $\mathrm{Eq}(R, CR)$ *are decidable.*

Proof. The families FT (finite-turn dcfls) and R (real-time dcfls) fulfill conditions (i), (ii), (iii) of Theorem III.3.

Remark III.1. As $D \subset CR$ (Theorem II.1), our method gives a new proof of the fact that one can decide the equivalence of two dpdas, one of which is finite-turn or real-time (Oyamaguchi *et al.*, 1986, 1987).

III.C. More about the Family of Church–Rosser Languages

In order to see that the inclusion $D \subset CR$ is strict, we give some examples of languages belonging to $CR - D$.

The first one is borrowed from Cochet (1975/1976) and Nivat (1970). It shows that some CR languages are not even context-free.

The second one is a CR language which is context-free but not deterministic context-free.

Example 1 (Cochet, 1975/1976; Nivat, 1970). $X = \{a, b, c\}$; S is the semi-Thue system $\{(ab, bba), (ca, caa)\}$:

(1) There are no overlapping right members in S. Hence S is trivially locally confluent and then it is Church–Rosser (by Lemmas 2.1 and 2.4 of Huet (1980)).

(2) $[cab] \overset{*}{\leftrightarrow}_S \cap cb^*a^* = \{cb^{2^n}a^n\}_{n \geqq 1}$ which is not context-free because it does not fulfill the classical iteration lemma for context-free languages (Bar-Hillel, Perles, and Shamir, 1961).

We remark that $\vdash_S = \vdash_T$, where T is the CR, controlled rewriting system: $T = \{(X^*, ab, bba), (X^*, ca, caa)\}$. Hence $[cab] \overset{*}{\leftrightarrow}_S$ belongs to the family CR but is not context-free.

Example 2. $X = \{a, b, x, y, z\}$, $\Sigma = X \cup \{\sigma, S, A, A'\}$; S is the semi-Thue system consisting of the rules

$$\sigma \to Sx; \qquad \sigma \to Sy; \qquad S \to bSA; \qquad S \to bSA';$$

$$SA \to zzA; \qquad SA' \to zzA'; \qquad AA \to aaaA; \qquad A'A' \to aaA';$$

$$Ax \to aaax; \qquad A'y \to aay.$$

Fact 1. *$\vdash_S$ is Church–Rosser.*

Proof. The arguments used in point (2) of Example 1 are still available here. ∎

Fact 2. $[\sigma] \overset{*}{\leftrightarrow}_S \cap X^* = \{b^n z^2 a^{3n} x\}_{n \geqq 1} \cup \{b^n z^2 a^{2n} y\}_{n \geqq 1}$.

Corollary. *$[\sigma] \overset{*}{\leftrightarrow}_S$ is not a deterministic* cfl.

Proof. If $[\sigma] \overset{*}{\leftrightarrow}_S$ was a deterministic cfl, then $[\sigma] \overset{*}{\leftrightarrow}_S \cap X^*$ would be a deterministic cfl too. But the arguments used in the proof of Theorem 4.1, p. 463 of Ginsburg and Greibach (1966) show that $\{b^n z^2 a^{3n} x\}_{n \geqq 1} \cup \{b^n z^2 a^{2n} y\}_{n \geqq 1}$ is not deterministic.

FACT 3. *$[\sigma] \overset{*}{\longleftrightarrow}_S$ is context-free.*

Proof. Let us look at the semi-Thue system $\tilde{S}$ obtained by replacing each pair $(u, v) \in S$ by $(\tilde{u}, \tilde{v})$, where $\tilde{f}$ denotes the reversal image of f. The direct reduction $\vdash_{\tilde{S}}$ is equal to $\vdash_T$, where T is the following controlled rewriting system:

$$T = \{(\Sigma^*, \sigma, xS), (\Sigma^*, \sigma, yS), (\Sigma^*, S, ASb), (\Sigma^*, S, A'Sb),$$
$$(\Sigma^* A, S, zz), (\Sigma^* A', S, zz), (\Sigma^* A, A, aaa),$$
$$(\Sigma^* A', A', aa), (\Sigma^* x, A, aaa), (\Sigma^* y, A', aa)\}.$$

T is Church–Rosser (by Fact 1) and monadic; that is

$$T = \{(R_i, u_i, v_i)\}_{i \in [1,n]}, \text{ where for each } i \in [1, n],\ |u_i| \leqq 1, \text{ and } |u_i| < |v_i|.$$

Hence $[\sigma] \overset{*}{\longleftrightarrow}_T$ is deterministic context-free (it is proved in (Chottin, 1982) that every class modulo the equivalence generated by a Church–Rosser and basic controlled rewriting system is a dcfl. "Basic" is a generalisation of "monadic." One can also adapt the proof given in O'Dunlaing (1983) that for every regular set R and monadic, rational, Church–Rosser, Thue system T, $[R] \overset{*}{\longleftrightarrow}_T$ is a dcfl).

But $L = [\sigma] \overset{*}{\longleftrightarrow}_S = [\tilde{\sigma}] \overset{*}{\longleftrightarrow}_T$, hence L is context-free.

This example gives also a positive answer to the question asked in Boasson and Sénizergues (1985) which is : "Does there exist a finite Church–Rosser Thue system S such that one class modulo (S) would be context-free but not NTS." Since L is not a dcfl it cannot be NTS.

IV. REMARKS AND CONCLUSION

The family *NTS* contains *RAT*, is effectively closed under intersection with regular sets and has a decidable equivalence problem, but we still do not know if it is effectively closed under inverse homomorphism. It is conjectured in Boasson and Sénizergues (1985) than *NTS* is closed under Φ^{-1} and proved in Autebert *et al.*, (1984-1) that a subclass of *NTS*, the family of nestsets, has all its inverse homomorphic images in *NTS*.

In Definition II.1 we consider controlled rewriting systems S which are, by definition, strictly length decreasing: if $f \vdash g$ then $|f| > |g|$. We could consider more general controlled rewriting systems by replacing the condition "$|u_i| < |v_i|$" in Definition II.1 by "$\vdash_S$ is a noetherian relation" (as defined in Huet, 1980). We recall it means that there is no infinite sequence of words $(f_i)_{i \geqq 0}$ such that, $\forall i \geqq 0$, $f_i \vdash_S f_{i+1}$.

Let us call such systems "noetherian controlled rewriting systems" and denote by NCR the family of languages defined as

$$[R] \overset{*}{\underset{S}{\longleftrightarrow}}$$

for a noetherian and Church–Rosser controlled rewriting system S and a regular set $R \subset \mathrm{Irr}(S)$.

Theorems III.1 and III.3 remain true for this extended family (because the proofs given above are valid in this extended context).

Y. Cochet (1975/1976) noticed that languages in CR must be context-sensitive because they are recognized by linear bounded Turing machines. On the other hand NCR contains noncontext-sensitive languages (see in (Bauer and Otto, 1984), a method for constructing such languages).

By replacing the notion of "finite set of axioms $A \subset V$" in Definition II.3 by

regular set of axioms $A \subset (X \cup V)^*$ such that $\hat{L}(G, A) = [A] \overset{*}{\longleftrightarrow}_G$,

we get the definition of a generalized NTS language. This family fulfills all the interesting properties of NTS and is closed under complementation, right-quotient, and left-quotient by regular sets (Sénizergues, 1981). Theorem III.1 remains true if we replace NTS by $GNTS$ (because all the lemmas preparing Theorem III.1 remain true in this context).

PROPOSITION IV.1. *The equivalence problem between*

— *a language M defined as $M = [R] \overset{*}{\longleftrightarrow}_S$, where R is a regular set and S is a regular, monadic, Church–Rosser semi-Thue system*

— *and a language L in CR is a decidable problem.*

Sketch of Proof. Proposition III.1 still applies. M is a dcfl (O'Dunlaing, 1983); hence conditions (1), (2) are decidable.

Proposition II.1 is also true for the languages $\Delta(A)$ and $\delta(A)$ defined by $\vdash_S$ and $\mathrm{Irr}(S)$ instead of $\vdash_G$ and $\mathrm{Irr}(G)$.

It is then sufficient to show that $\mathrm{Synt}_X(R, M)$ is computable for every R regular set included in $\mathrm{Irr}(S)$. We define $C'(R, m) = \{u \# v \mid u \in R, v \in \mathrm{Irr}(S)$ and $umv \in M\}$. Lemma III.3 remains true for $C'(R, m)$.

A full proof of this can be found in pp. 183–211 of Sénizergues (1987). ∎

ACKNOWLEDGMENTS

I thank J. M. Autebert and L. Boasson for many discussions about a first draft of this paper. I thank B. Courcelle for discussions and positive criticism about a second draft.

RECEIVED February 4, 1986; ACCEPTED December 23, 1986

References

Autebert, J. M., Beauquier, J., Boasson, L., and Sénizergues, G. (1984-1), Remarques sur les langages de parentheses, *Theoret. Comput. Sci.* **31**, 337–349.

Autebert, J. M., Boasson, L., Sénizergues, G. (1984-2), Langages de parentheses, langages N.T.S. et morphismes inverses, *RAIRO Inform. Theor.* **18**.

Bar-Hillel, Y., Perles, M., and Shamir, E. (1961), On formal properties of simple phrase structure grammars, *Z. Phonet. Sprachwissensch. Kommun.* **14**, 143–172.

Bauer, G., and Otto, F. (1984), Finite complete rewriting systems and the complexity of the word problem, *Acta Inform.*, 522–540.

Beeri, C. (1976), An improvement on Valiant's decision procedure for equivalence of deterministic pushdown automata, *Theoret. Comput. Sci.* **3**, 305–320.

Benois, M. (1969), Parties rationelles du groupe libre, *C.R. Acad. Sci. Paris Sér. A* **269**, 1188–1190.

Boasson, L. (1980), Dérivations et réductions dans les grammaires algébriques, *in* "Proceedings, ICALP," Lecture Notes in Comput. Sci., Vol. 85, p. 109–118, Springer-Verlag, New York/Berlin.

Boasson, L., and Sénizergues, G. (1985), N.T.S. languages are congruential and deterministic, *J. Comput. System Sci.* **31**, 332–342.

Book, R. V., Jantzen, M., and Wrathall, C. (1982), Monadic Thue systems, *Theoret. Comput. Sci.* **19**, 231–251.

Butzbach, P. (1978), Sur l'équivalence des grammaires simples, "Actes des journées d'informatique théorique de Bonascre" (1973) (J. P. Crestin et M. Nivat, Eds.), pp. 223–245, ENSTA.

Chottin, L. (1979), Strict deterministic languages and controlled rewriting systems, *in* "Proceedings, ICALP," 1979, Lecture Notes in Comput. Sci. Vol. 71, Springer-Verlag, New York/Berlin.

Chottin, L. (1982), Langages algébriques et systèmes de réécriture rationnels, *RAIRO Inform. Théor.* **16**, No. 2.

Cochet, Y. (1975/1976), Langages définis par des congruences, *in* "Groupe d'étude d'algèbre" (M. P. Malliavin, Ed.), 1st ed.

Frougny, C. (1980), "Une famille de langages congruentiels: les langages à non-terminaux séparés," Thèse de 3ème de Maths, Paris 7.

Ginsburg, S., and Greibach, S. (1966), Deterministic context-free languages, *Inform. and Control* **9**, 620–648.

Huet, G. (1980), Confluent reductions, abstract properties, and applications to term rewriting systems, *J. Assoc. Comput. Mach.* **27**, 797–821.

Nivat, M. (1970), On some families of languages related to the Dyck language, *in* "2nd Symposium on Theory of Computing," pp. 221–225.

Nivat, M., and Benois, M. (1971/1972), Congruences parfaites et quasi-parfaites, *in* "Séminaire Dubreil, 25ème année."

O'Dunlaing, C. (1983), Infinite regular Thue systems, *Theoret. Comput. Sci.* **23**, 339–345.

Oyamaguchi, M. (1986), personnal communication.

Oyamaguchi, M. (1987), The equivalence problem for real-time dpda's, *J. Assoc. Comput. Mach.* **34**, No. 3, 731–760.

Oyamaguchi, M., Inagaki, Y., and Honda, N. (1981), The equivalence problem for two dpda's one of which is a finite-turn or one-counter machine, *J. Comput. System Sci.* **23**, No. 3.

Oyamaguchi, M., Inagaki, Y., and Honda, N. (1980), On the equivalence problem for two dpda's one of which is real-time, *in* "Proceedings IFIP Congress 80, Tokyo," pp. 53–57.

Sakarovitch, J. (1979), "Syntaxe des langages de Chomsky Essai sur le déterminisme," Thèse d'état de Maths, Paris 7.

SÉNIZERGUES, G. (1981), "Décidabilité de l'équivalence des grammaires N.T.S.," Thèse de 3ème cycle de Maths, Paris 7.

SÉNIZERGUES, G. (1985), The equivalence and inclusion problems for NTS languages, *J. Comput. System Sci.* **31**, No. 3, 303–331.

SÉNIZERGUES, G. (in press 1), A characterisation of deterministic context-free languages by means of right-congruences, *Theoret. Comput. Sci.*

SÉNIZERGUES, G. (in press 2), Some decision problems about controlled rewriting systems, *Theoret. Comput. Sci.*

SÉNIZERGUES, G. (1987), "Sur la description des langages algébriques déterministes par des systèmes de réécriture confluents," Thèse d'état, université Paris 7, Juillet 1987; Rapport LITP 88-39.

VALIANT, L. G. (1974). The equivalence problem for deterministic finite-turn pushdown automata, *Inform. and Control* **25**, 123–133.

VALIANT, L. G., AND PATERSON, M. S. (1975), Deterministic one-counter automata, *J. Comput. System Sci.* **10**, 340–350.